**An Approximate Dynamic Programming Model For MEDEVAC Dispatching**

THESIS

MARCH 2015

Aaron J. Rettke, Captain, USA

AFIT-ENS-MS-15-M-115

**DEPARTMENT OF THE AIR FORCE**
**AIR UNIVERSITY**

***AIR FORCE INSTITUTE OF TECHNOLOGY***

**Wright-Patterson Air Force Base, Ohio**

AFIT-ENS-MS-15-M-115

# AN APPROXIMATE DYNAMIC PROGRAMMING MODEL FOR MEDEVAC DISPATCHING

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Aaron J. Rettke, BS, MS

Captain, USA

MARCH 2015

AFIT-ENS-MS-15-M-115

# AN APPROXIMATE DYNAMIC PROGRAMMING MODEL FOR MEDEVAC DISPATCHING

THESIS

Aaron J. Rettke, BS, MS
Captain, USA

Committee Membership:

Lt Col Matthew J. Robbins, PhD
Chair

LTC Brian J. Lunday, PhD
Member

AFIT-ENS-MS-15-M-115

## Abstract

We develop a Markov decision process (MDP) model to examine military medical evacuation (MEDEVAC) dispatch policies. To solve the MDP, we apply an approximate dynamic programming (ADP) technique. The problem of deciding which aeromedical asset to dispatch to which service request is complicated by the service locations and the priority class of each casualty event. We assume requests for MEDEVAC arrive sequentially, with the location and the priority of each casualty known upon initiation of the request. The proposed model finds a high quality dispatching policy which outperforms the traditional myopic policy of sending the nearest available unit. Utility is gained by servicing casualties based on both their priority and the actual time until a casualty arrives at a medical treatment facility (MTF). The model is solved using approximate policy iteration (API) and least squares temporal difference (LSTD). Computational examples are used to investigate dispatch policies for a scenario set in northern Syria. Results indicate that a myopic policy is not always the best policy to use for quickly dispatching MEDEVAC units, and insight is gained into the value of specific MEDEVAC locations.

Key words: Emergency Medical Dispatch, medical evacuation (MEDEVAC), Markov decision processes, approximate dynamic programming, approximate policy iteration, least squares temporal difference

*I would like to thank my family for their limitless support. I owe my wife immeasurably as none of my accomplishments would have been possible without her patience, compassion, and encouragement. Words can not express my love and appreciation for my two lovely children who, without fail, remind me to smile every day.*

*This research is dedicated to the Soldiers, past present and future, who gave their lives in support of our country; may this research continue in order to provide the best medical evacuation system possible for our Soldiers.*

# Acknowledgements

I would like to express my deep gratitude to Dr. Robbins for his mentorship and counsel. Without his supervision and guidance this thesis would not have been possible. In addition, I would like to thank my committee member, Dr. Lunday. Their expertise and professional dedication greatly assisted in the development of this study.

Aaron J. Rettke

# Table of Contents

# List of Figures

# List of Tables

AN APPROXIMATE DYNAMIC PROGRAMMING MODEL
FOR MEDEVAC DISPATCHING

# I. Introduction

The United States military uses two classification categories when evacuating injured soldiers and civilians from the battlefield. These are medical evacuation (MEDEVAC) and casualty evacuation (CASEVAC). The primary and preferred method is MEDEVAC, which constitutes dedicated medical personnel on board the vehicle that is transporting casualties. The second method, often used as a contingency, is CASEVAC, in which there are no dedicated medical personnel on board to attend to a casualty event while in transit to a medical treatment facility (MTF) (Department of the Army, 2007). Any type of vehicle may be used to conduct MEDEVAC and CASEVAC operations.

When a request for a MEDEVAC occurs, there are three categories of evacuation precedence (Department of the Army, 2007):

**Priority I**, *Urgent*: Assigned to emergency cases that should be evacuated as soon as possible and within a maximum of 1 hour in order to save life, limb, or eyesight, to prevent complications of serious illness, or to avoid permanent disability.

**Priority II**, *Priority*: Assigned to sick and wounded personnel requiring prompt medical care. This precedence is used when the individual should be evacuated within 4 hours or when an individual's medical condition could deteriorate to such a degree that he or she will become an URGENT precedence, or whose requirements for special treatment are not available locally, or who will suffer unnecessary pain or disability.

**Priority III**, *Routine*: Assigned to sick and wounded personnel requiring evacuation but whose condition is not expected to deteriorate significantly. The sick and wounded in this category should be evacuated within 24 hours.

Aerial MEDEVAC operations involve the use of dedicated helicopters, specifically the UH-60 A/L Blackhawk. Helicopters are uniquely suited for MEDEVAC operations because they are able to travel faster, further, and access terrain which is not accessible to ground vehicles. The ability to simultaneously treat and quickly transport casualties from the point of injury (POI) to an MTF greatly increases the chance of survival for casualties. During the Korean War, the United States military experienced the first large-scale use of helicopters to remove casualties from the battlefield. Currently, survivability of injuries on the battlefield are at a historic high; 90% of all casualties survive, compared to 84% in Vietnam and 80% in World War II (Eastridge *et al.*, 2012). This improvement is attributed primarily to the speed in which casualties are able to receive proper medical attention. A quote from United States Army Surgeon, Major General Neel Spurgeon:

> "Getting the casualty and the physician together as soon as possible is the keystone of the practice of combat medicine...(Spurgeon, 1991)"

Recent conflicts have exhibited a shift from traditional force-on-force engagements to counterinsurgency operations (COIN). With COIN, units are typically much smaller and more geographically dispersed, causing a greater dispersion of critical resources. As the area of the battlefield increases, helicopters provide a lifeline to these soldiers, allowing them to operate further away from bases while still able to receive aerial support. From the inception of Operation Enduring Freedom (OEF) in

2001 until 2008, more than 3200 casualties have been transported using MEDEVAC (Hartenstein, 2008).

When utilizing MEDEVAC assets, three different aspects need to be considered: location, dispatching, and redeployment. The location of MEDEVAC assets is a balance between maximizing coverage and minimizing response time. Placement may be further constrained by force protection and maintenance necessities. Dispatching of MEDEVAC units is often conducted using a myopic policy in which the nearest MEDEVAC is launched to a POI regardless of its priority. The third aspect, dynamic redeployment of ambulances, is alse possible. However, communication and crew limitations often make this problematic and is not typically performed.

In this thesis we consider the MEDEVAC dispatching problem in which a dispatching authority must decide which MEDEVAC to send in response to a request for MEDEVAC. Redeployment is not considered. A Markov decision process (MDP) is contructed to model this MEDEVAC dispatching problem. We utilize an Approximate Dynamic Programming (ADP) approach to obtain high quality solutions to the problem. The proposed ADP algorithm utilizes Least Squares Temporal Difference (LSTD) policy evaluation within an Approximate Policy Iteration (API) framework. Bellman error minimization is applied in the policy improvement phase to obtain improved policies. To demonstrate the applicability of the model to the medical planning process, we present a notional scenario involving the allied defense of northern Syria in response to aggression by Islamic State (IS) militants.

This thesis is organized as follows. In Chapter 2 we review related research in the contemporary literature. In Chapter 3 we present the MDP formulation of the problem and ADP algorithm to solve it. Chapter 4 contains the computational scenario and testing results. In Chapter 5 we present the findings and conclusions.

# II. Literature Review

Our literature review focuses on two research areas. First, we examine research in emergency medical systems (EMS) and MEDEVAC in order to inform the model of the MEDEVAC dispatching problem. Next, research is focused on the field of ADP to inform the development of the solution methodology.

## 2.1 EMS and MEDEVAC

The nature of MEDEVAC operations shares many inherent similarities with EMS. Decisions need to be made quickly regarding which unit will serve a specific casualty event. We initially examine research into EMS optimization, which can be traced back to the late 1960's and early 1970's with papers on optimally locating EMS units. Church & ReVelle (1974) examine the maximal coverage location problem (MCLP), ensuring there is an ambulance within a specific distance or time from a POI.

ReVelle & Hogan (1989) introduce a MCLP which builds upon the basic maximal covering problem by ensuring there is always an ambulance available within a predetermined length of time in the event of another ambulance being unavailable. Alsalloum & Rand (2006) extend the approach used by ReVelle & Hogan (1989); they determine the minimum number of vehicles required to cover the largest possible area given a set of constraints.

Batta *et al.* (1989) extend the MCLP by examining not only busy probabilities for EMS units, but also the queuing of calls via the Hypercube model. Silva & Serra (2008) expand upon covering problems by incorporating queuing theory as well as establishing different patient priority levels.

Although locating assets is a critical aspect of EMS and MEDEVAC systems, the limitations for MEDEVAC placement are beyond the scope of this thesis. We examine how to optimize dispatching. Initial research by Carter *et al.* (1972) found that dispatching the nearest unit to a service did not always produce the lowest average response time. This result is very important, as most EMS and MEDEVAC systems operate using a myopic policy of sending the nearest unit to any patient.

More recent research by McLay & Mayorga (2010) utilize real-world EMS data to examine optimal dispatching policies. In their model, McLay & Mayorga consider fixed bases for responding units and measured response time thresholds (RTT) as a measure for patient survivability. McLay & Mayorga obtain policies that decrease RTT for high priority patients compared to a myopic policy.

Bandara *et al.* (2012) build upon the work of McLay & Mayorga (2010) using patient survivability instead of RTT. This measure more closely mirrors patient outcomes, provide more insight into the significance of any improvement. Mayorga *et al.* (2013) continue to add depth to dispatching research by designing a constructive heuristic to identify response districts for EMS units. They compare system performance between respective policies of allowing units to service other districts and forcing them to stay in their own district. These results are compared against myopic and heuristic policies, and it is shown that all three policies perform better than the myopic policy.

In order to optimize the MEDEVAC system, an objective function must be determined. Erkut *et al.* (2008) question the terms "coverage" and "performance" for an EMS system. They propose the use of a monotonically decreasing function over time for the probability of patient survivability as a measure of performance. Subsequent work by Bandara *et al.* (2012) and Grannan *et al.* (2014) similarly use survivability functions instead of the traditional RTT as a performance measure for their model. A problem with the survivability function is the difficulty finding empirical evidence to support a particular functional form. Eastridge *et al.* (2012) provide extensive statistics about combat deaths, but the response times are not known, preventing the creation of an appropriate survivability function. Although Feero *et al.* (1995) examine EMS response times relative to trauma survivability, their work is limited to response times under eight minutes. MEDEVAC units typically need to travel significantly further than EMS units, and so response times are significantly longer. The goal of U.S. MEDEVAC support is to respond within 60 minutes from notification to drop-off of the patient at a MTF (Garrett, 2013). The one-hour response is sometimes referred to as the "Golden Hour". Lerner & Moscati (2001) examine the "Golden Hour," which is mentioned in numerous trauma articles. However, its exact origins and any quantifiable measure are not reported. We assume an exponentially decreasing function to model patient outcome.

Research specifically into military MEDEVAC systems has been conducted recently by Zeto *et al.* (2006), Fulton *et al.* (2010), Bastian (2010), Grannan *et al.* (2014) and Keneally *et al.* (2014). Much of the research in military MEDEVAC concerns optimal emplacement of assets. Zeto *et al.* use a maxi-min goal programming approach adapted from Alsalloum & Rand (2006) in order to maximize coverage and minimize the response times for MEDEVAC units in the Afghanistan theater. Fulton *et al.* (2010) develop a stochastic model to determine where to emplace multiple medical assets such as air and ground MEDEVAC units. Bastian (2010) develops a model focusing on the performance measure of the UH-60 A/L combined with austere and hostile conditions in which it operates. Using a combination of goal programming and stochastic optimization, Bastian (2010) seeks to optimally emplace MEDEVAC assets in Afghanistan.

Keneally *et al.* (2014) develop an MDP to examine MEDEVAC dispatch policy in regional command south (RC-S) of the OEF theater. They use two priority classifications, a reward function based on RTT, and Monte Carlo simulation with a Hawkes process for casualty generation. The find that the myopic policies do not always lead to optimal performance.

## 2.2 ADP

We also consider an MDP model. However, the curse of dimensionality prevents us from applying the MDP solution techniques used by Keneally *et al.* (2014). We turn our attention to ADP which has emerged as a technique to solve large or complex problems for making sequential decisions under uncertainty. Powell (2012) provides a broad overview of ADP and its origins from different research communities. He found many communities use similar methods or algorithms to battle the curse of dimensionality that many systems or problems face. The three curses of dimensionality that affect systems concern the state space, the outcome space, and the action space. Powell presents four classes of policies, which refer to the mapping of a state to an action, as being myopic cost function approximations, look-ahead policies, policy function approximations, and value function approximations. ADP approaches also incorporate hybrid policies, i.e. combining two or more classes. Researchers seeking such policies are able to employ ADP techniques to determine such policies and solve otherwise intractable MDPs. We use value function approximation when solving the dispatching problem.

Maxwell *et al.* (2010) present an ADP approach to examine the potential of dynamically redeploying ambulances to maximize the number of patients that are served within an RTT. They set the framework of their system using an MDP with a similar state space to the MEDEVAC problem. At each event occurrence in their system, such as a new call arrival or an ambulance bringing a patient to the hospital, a decision is made to redeploy any available units to better cover the service area. They obtain a redeployment policy for this this high dimensional system using approximate value iteration. They are able to demonstrate a significant improvement in EMS response at two metropolitan cities using dynamic redeployment versus a static myopic policy.

Although there is much potential in dynamic routing and positioning of MEDEVAC units, the complexity of helicopter operations limits the ease and practicality of dynamic routing. This thesis assumes all MEDEVAC units return to their base upon service completion.

Bradtke & Barto (1996) introduce three temporal difference algorithms: normalized temporal difference (NTD), recursive least squares temporal difference (RLSTD), and LSTD. Temporal difference learning allows the system to learn the expected value for a state-action pair. They prove all three converge to optimality when used with a functional approximator. LSTD and RLSTD are able to extract more information from each observation which in turn allows each algorithm to converge to optimality faster than NTD. LSTD and RLSTD do require more computational for time each observation; however, this is offset by its faster convergence rate.

Lagoudakis & Parr (2003) introduce a least squares policy iteration (LSPI) algorithm which builds upon LSTD. Policy iteration (PI), first introduced by Howard (1960), is a simple two-step iterative algorithm for stationary policies. First, the policy is evaluated to determine its value. Second, we attempt to improve the policy by finding a policy variant which is monotonically increasing in value. API introduces approximations to represent the value function and policy in order to make these problems computationally tractable. Lagoudakis & Parr (2003) compare the performance of LSPI to other reinforcement learning algorithms. They find LSPI performs significantly better than Q-learning and is computationally faster.

# III. Methodology

## 3.1 MDP Formulation

When a casualty occurs and a MEDEVAC request is received, a decision must be made quickly regarding which MEDEVAC asset to dispatch. Any delays in decision making affect casualties' survivability. Thus, it is critical to quickly and accurately determine a high quality solution. The stochastic elements in the model are depicted in Figure 1.

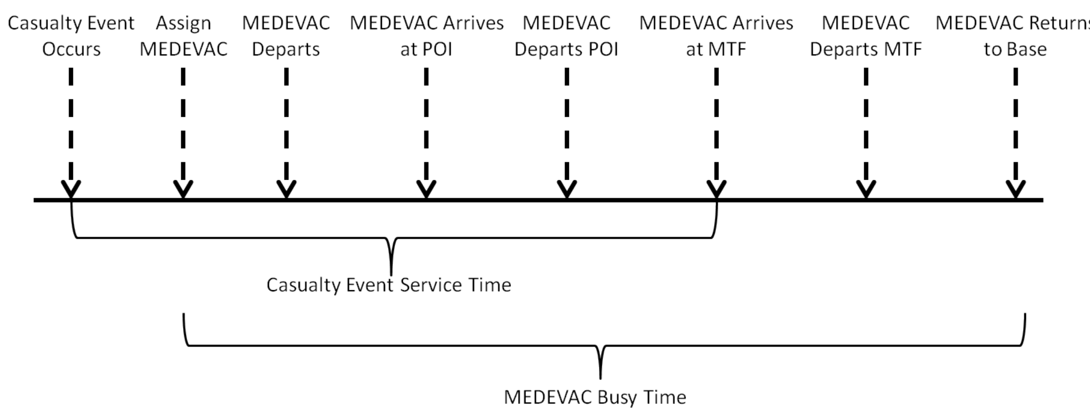



**Figure 1. MEDEVAC Dispatch Timeline**

We consider three casualty event categories: *urgent*, *priority*, and *routine*. In the model, casualties are generated from clusters using a Poisson Hawkes process with rate $\lambda$. Response times are independent for each casualty event and priority classification. A monotonically decreasing function based on service time and casualty category is used as the reward function. We incorporate queuing in the model to allow multiple casualty events to occur and the decision to wait before launching a MEDEVAC. The decision to wait can be advantageous if a low priority casualty event has occurred while other MEDEVAC units are busy; it may be better to remain on standby until other MEDEVAC units become available in case a high priority casualty event occurs.

All casualties are evacuated to the nearest MTF. After a group of casualties is dropped off, MEDEVAC units return to their originating base.

To model the system, we base the framework on the work of Maxwell *et al.* (2010). We establish the state space for the MDP model as a tuple with a time state, an event state, a MEDEVAC status vector, and a queue status vector. We represent the system with the tuple $s = (\tau, \epsilon, M, Q)$, wherein $\tau$ corresponds to the current system time and $\epsilon$ corresponds to the current event. The main system components are $M = (m_1, m_2, ..., m_a)$ and $Q = (q_1, q_2, ..., q_b)$, where $m_i$ contains information about the $i^{th}$ MEDEVAC, $a$ represents the maximum number of MEDEVACs in the system, $q_j$ contains information about the $j$th casualty event in the queue, and $b$ represents the maximum number of casualty events allowed in the queue. The state of MEDEVAC $i$ is given as a tuple $m_i = (\sigma_i, d_i, t_i)$, wherein $\sigma_i$ is the status of the MEDEVAC, $d_i$ is the expected time to complete the current movement, and $t_i$ is the starting time of any MEDEVAC movement. Once a MEDEVAC is launched to service a casualty event, the queue status is updated. After a MEDEVAC drops off a group of casualties, it returns to base where it then becomes available to launch. The status of the MEDEVAC, $\sigma_i$, can be "idle", "enroute to a casualty event", "at POI", "enroute to MTF", or "returning to base". If the MEDEVAC is not idle, $t_i$ corresponds to the starting time of the movement; otherwise, $t_i$ represents the time of the current event cycle. The state of casualty event $j$ in the queue is $q_j = (\delta_j, l_j, \zeta_j, \eta_j)$, where $\delta_j$ is the status of the casualty event in the $j^{th}$ position, $l_j$ is the location of the casualty event, $\zeta_j$ is the time the casualty event arrived in the system, and $\eta_j$ is the priority of the casualty event.

Events are triggered by changes in the status of a MEDEVAC or an arrival of a call. The event list is given in Table 1. The model assumes MEDEVAC dispatch decisions only occur when an a event occurs. Although it is possible to reroute a MEDEVAC

**Table 1. Event List**

| $\epsilon$, event list: |
|---|
| Call arrives and is placed in the $j$th position |
| MEDEVAC $i$ departs for call $j$ at casualty event |
| MEDEVAC $i$ arrives at caualty event for call $j$ |
| MEDEVAC $i$ leaves call $j$ casualty event for MTF |
| MEDEVAC $i$ delivers call $j$ at MTF |
| MEDEVAC $i$ departs MTF to return to base |
| MEDEVAC $i$ arrives at base |

mid-flight, delay and confusion in communication can cause large problems and this practice is not typical in combat operations.

Let $\mathscr{M}(s) = \{i : \sigma_i = \text{“}idle\text{”}\}$ denote the set of MEDEVACs available for dispatching when the system is in state $s$.

Let $\mathscr{Q}(s) = \{j : \delta_j = \text{“}idle\text{”}\}$ denote the set of casualty events awaiting service by MEDEVAC when the system is in state $s$.

To capture dispatching decisions we let $x_{ij}(s)$=1 if MEDEVAC $i$ is deployed to casualty event $j$ when the system is in state $s$, and 0 otherwise. The set of feasible decisions can be written as:

$$\mathscr{X}(s) = \left\{ x(s) \in \{0,1\}^{|\mathscr{M}(s)| \times |\mathscr{Q}(s)|} : \sum_{i \in \mathscr{M}} \sum_{j \in \mathscr{Q}} x_{ij}(s) \leq 1 \right\}. \quad (1)$$

The trajectory of the system is denoted in the form $\{(s_k, x_k) : k = 1, 2, ...\}$ where $s_k$ is the state of the system, and $x_k$ is the decision after the $k$th event has occurred.

To capture the dynamics of the system, the following symbology is used: $s_{k+1} = f(s_k, x_k, \omega(s_k, x_k))$, where $\omega(s_k, x_k)$ is a random element of an appropriate space representation of the stochastic process related to casualty event arrivals and delays, and where $f$ is the transfer function.

The reward function is determined by the value of being in the pre-decision state $s_k$. The reward, $r(s_k, x_k, s_{k+1})$ is shown in Equation 2.

$$r(s_k, x_k, s_{k+1}) = \begin{cases} \Delta_1 \cdot 0.99^{(t_{k+1}-\zeta_j)/60} & \text{if } \eta_j = \text{“}\textit{urgent}\text{” and } \epsilon(s_{k+1}) \text{ is such} \\ & \text{that casualty event } j \text{ was dropped off at a MTF,} \\ \Delta_2 \cdot 0.99^{(t_{k+1}-\zeta_j)/240} & \text{if } \eta_j = \text{“}\textit{priority}\text{” and } \epsilon(s_{k+1}) \text{ is such} \\ & \text{that casualty event } j \text{ was dropped off at a MTF,} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Tunable parameters $\Delta_1$ and $\Delta_2$ are used to model the value trade-off between different priority casualty events. We normalize the time the casualty event has been in the system based on requirements outlined in Department of the Army (2007). The reward obtained by servicing higher priority casualties decays quickly; however, servicing higher priority casualties obtains higher utility and they are therefore often served first. For lower priority casualties, the decay allows enough time to wait before being forced to launch immediately, which increases the flexibility of the decision maker.

A policy $\pi(s) \in \mathscr{X}(s)$ maps the state space to the action space. In this form, $\pi(s)$ is the action taken when the system is in state $s$. Following policy $\pi$ the state trajectory of the system $\{s_k^\pi : k = 1, 2, ...\}$ evolves according to $s_{k+1}^\pi = f(s_k^\pi, \pi(s_k^\pi), \omega(s_k^\pi, \pi(s_k^\pi)))$. The objective function is given by Equation 3,

$$J^\pi(s) = \mathbb{E}\left[\sum_{k=1}^{\infty} \gamma^{\tau(s_k^\pi)} r(s_k^\pi, \pi(s_k^\pi), s_{k+1}^\pi) | s_1^\pi = s\right], \tag{3}$$

where $\gamma \in [0, 1)$ is the discount factor and $\tau(s_k^\pi)$ is the time at which the system is in state $s_k^\pi$. The optimal policy $\pi^*$ maximizes the expected total discounted reward and satisfies the following optimality Equation 4.

$$J(s) = \max_{x \in \mathscr{X}(s)} \left\{ \mathbb{E}\left[ r(s, x, f(s, x, \omega(s, x))) + \gamma^{\tau(f(s,x,\omega(s,x))) - \tau(s)} J(f(s, x, \omega(s, x))) \right] \right\} \tag{4}$$

The number of possible values created by this equation is uncountable. Even if the state space was discrete, it would be computationally intractable. In the next section we propose an ADP approach to obtain an approximation of $J(s)$. Utilizing the value function approximation for $J(s)$, high quality policies are constructed and compared to a simple myopic policy, which is often employed in practice.

### 3.2 ADP Formulation

We use LSTD combined with API (as discussed by Scott *et al.* (2014)) to approximate an optimal solution. API is very similar to policy iteration, which is used to solve classical MDPs. To obtain an optimal policy $\pi^*$ we need to solve Equation 4. To solve the problem, we construct an approximation of the value function. We employ a modified version of Bellman's equation that uses *post-decision state* variables. The post-decision state $s_k^x$ refers to the state of the system after being in state $s_k$ and upon taking action $x_{ij}$. The post-decision state variable provides tremendous computational advantages as its use eliminates the embedded expectation within the Bellman equation. The value of being in state $s_k^x$ immediately after a decision is made is denoted by $J^x(s_k^x)$. The relationship between $J(s)$ and $J^x(s_k^x)$ is defined as:

$$J^x(s_k^x) \stackrel{\text{def}}{=} \mathbb{E}[J(s_{k+1}) \mid s_k^x]. \tag{5}$$

Bellman's equation in the post-decision state becomes:

$$J^x_{k-1}(s^x_{k-1}) = \max_x \{\mathbb{E}[r(s_k, x) + \gamma J^x_k(s^x_k)) \mid s^x_{k-1}]\}. \tag{6}$$

Despite the reduction in dimensionality with the use of the post-decision state, Equation 6 is still computationally intractable for the model. We proceed by developing a set of basis functions to approximate $J^x(s^x_k)$. Equation 7 shows the formulation we use in the model.

$$\overline{J}^x(s^x_k) \stackrel{\text{def}}{=} \sum_{f \in \mathscr{F}} \theta_f \phi_f(s^x_k) = \theta^\top \phi(s^x_k), \tag{7}$$

where $\phi(s^x_k)$ is a column vector with elements $\{\phi_f(s^x_k)\}_{f \in \mathscr{F}}$ and $\theta$ is a column vector of basis function weights. By substituting the value function approximation Equation 7 into the Bellman equation using the post-decision state variable (Equation 6), we obtain the following expression:

$$\theta^\top \phi(s^x_{k-1}) = \mathbb{E}[r(s_k, X^\pi(s_k|\theta)) + \gamma \theta^\top \phi(s^x_k) \mid s^x_{k-1}]. \tag{8}$$

Since the above equation is an approximation of the multidimensional model, a linear model may not provide a fixed solution. However, we are still able to use this representation as the foundation of the Least Squares Aproximate Policy Iteration (LSAPI) algorithm. We find the policy decision, $x^\pi_{ij}(s_k|\theta)$, by solving Equation 9

$$X^\pi(s_k|\theta) = \arg\max_x [r(s_k, x) + \gamma \theta^\top \phi(s^x_k)]. \tag{9}$$

The API algorithm shown in Table 2 was introduced by Bradtke & Barto (1996) and modified by Ma & Powell (2010). Starting with an initial $\theta$ for the base policy, we then step into the policy improvement loop. To evaluate the performance of the policy, the post-decision state is randomly sampled and the value $\phi(S^x_{k-1,h})$ is recorded. Next, we simulate one event forward and determine the optimal decision based on Equation 9, recording the associated reward, $r(S_{k,h})$, and basis function values of the post-decision state, $\phi(S^x_{k,h})$. After completing the sampling of the post-decision state space, we evaluate the performance of the current policy. We also introduce a harmonic step-size rule, as indicated in Equation (10), to smooth $\theta$. Smoothing is required because we are sampling the state space to approximate the Bellman Equation 4. Were Equation 4 computationally tractable, the model could be solved using traditional value iteration, and a smoothing function would not be required (Powell (2009)). The parameters $\alpha$, $\mathscr{G}$, and $\mathscr{H}$ are all tunable, where $\alpha$ is a step size parameter, $\mathscr{G}$ is the number of policy improvement iterations completed and $\mathscr{H}$ is the number of policy evaluation iterations completed.

**Table 2. API Algorithm**

| Approximate Policy Iteration Algorithm |
|---|
| Step 1: Initialize $\theta$ |
| Step 2: for $g$=1 to $\mathscr{G}$ (Policy Improvement Loop) |
| Step 3: for $h$=1 to $\mathscr{H}$ (Policy Evaluation Loop) |
| Step 4: Simulate a random post-decision state, $S^x_{k-1,h}$. |
| Step 5: Record $\phi(S^x_{k-1,h})$. |
| Step 6: Simulate the state transition for the next event to get $S_{k,h}$. |
| Step 7: Determine the decision, $x_{i,j} = X^\pi(S_{k,h} \mid \theta)$ using Equation 9. |
| Step 8: Record $r(S_{k,h})$. |
| Step 9: Record $\phi(S^x_{k,h})$. |
| Step 10: End. |
| Step 11: Update $\theta$ using Equation 11 and step size Equation 10. |
| Step 12: End. |

$$\theta_{new} = \hat{\theta} \cdot \left( \frac{\alpha}{\alpha + g - 1} \right) + \theta \cdot \left( 1 - \frac{\alpha}{\alpha + g - 1} \right) \tag{10}$$

To evaluate the policy, we apply least squares regression. The post-decision states, $\phi(S^x_{k-1,h})$ and $\phi(S^x_{k,h})$, are regressed against the reward, $r(S_{k,h})$. We first establish the basis function matrices and reward vector. Let

$$\Phi_{k-1} \stackrel{\text{def}}{=} \begin{bmatrix} \phi(S^x_{k-1,1})^\top \\ \vdots \\ \phi(S^x_{k-1,PE})^\top \end{bmatrix}, \quad \Phi_k \stackrel{\text{def}}{=} \begin{bmatrix} \phi(S^x_{k,1})^\top \\ \vdots \\ \phi(S^x_{k,PE})^\top \end{bmatrix}, \quad r(S_k) \stackrel{\text{def}}{=} \begin{bmatrix} r(S_{k,1}) \\ \vdots \\ r(S_{k,PE}) \end{bmatrix}, \quad \bar{r}(S_k) \stackrel{\text{def}}{=} \begin{bmatrix} \bar{r}(S_{k,1}) \\ \vdots \\ \bar{r}(S_{k,PE}) \end{bmatrix},$$

where matrices $\Phi_{k-1}$ and $\Phi_k$ are rows of basis function evaluations of the sampled post-decision states, $r_k$ is the reward vector for the sampled events, and $\bar{r}_k$ is the expected reward. The difference between $r_k$ and $\bar{r}_k$ is what we will refer to as the Bellman error, which we seek to minimization. Lagoudakis & Parr (2003) show that, by minimizing Equation 11, an improved set of $\theta$ values can be attained,

$$\min ||r_k - \bar{r}_k||_2^2 \tag{11}$$

To capture the dynamics of this model, we need to properly represent the approximations of the model with basis functions. Basis functions can be very difficult to develop (Powell, 2012). We start by creating an indicator variable if MEDEVAC $i$ is available, as represented by Equation 12,

$$\phi_{1,i} = \begin{cases} 1 & \text{if } m_i = \text{“Idle”} \\ 0 & \text{otherwise .} \end{cases} \quad \forall\ i = 1, 2, ..., a \tag{12}$$

To capture the expected time until a MEDEVAC becomes available, the 2nd basis function is defined in Equation 13, where $d^r_{i,k}$ represents the expected time for MEDEVAC $i$ to return to base after event $k$, dropping off casualties at a MTF. This expected time is added to $d_i$ - $\tau$, which is the expected time for the MEDEVAC to complete its current movement.

$$\phi_{2,i} = \begin{cases} d_i - \tau + d^r_{i,k} & \text{if } m_i = \text{“g”} \\ 0 & \text{otherwise} \end{cases} \quad \forall\ i = 1, 2, ..., a \tag{13}$$

The next basis functions, $\phi_{3,i}, \phi_{4,i}$, and $\phi_{5,i}$, capture the status of all casualty events currently being served. Basis function $\phi_{3,i}$ represents the expected time from $\tau$ until MEDEVAC $i$ arrives at the nearest MTF with its assigned casualty event.

$$\phi_{3,i} = \begin{cases} d_i - \tau & \text{if } m_i = \text{“Serving casualty event”} j \\ 0 & \text{otherwise} \end{cases} \quad \forall\ i = 1, 2, ..., a \tag{14}$$

Basis function $\phi_{4,i}$ captures the total expected time (including launch delay) that a casualty event will be in the system if it was served by MEDEVAC $i$.

$$\phi_{4,i} = \begin{cases} \zeta_j - t_i + d_i & \text{if } m_i = \text{“Serving casualty event”} j \\ 0 & \text{otherwise} \end{cases} \quad \forall\ i = 1, 2, ..., a \tag{15}$$

Basis function $\phi_{5,i}$ captures the priority of the casualty being served by MEDEVAC $i$.

$$\phi_{5,i} = \begin{cases} \eta_j & \text{if } m_i = \text{“Serving casualty event”} j \\ 0 & \text{otherwise} \end{cases} \quad \forall\ i = 1, 2, ..., a \tag{16}$$

The final basis function calculates the expected time in system for casualty event $j$ if it is to be assigned MEDEVAC $i$.

$$\phi_{6,i,j} = d^*_{i,j} + \phi_{2,i} \quad \forall\ i = 1, 2, ..., a;\ j = 1, 2, ..., b, \tag{17}$$

where $d^*_{i,j}$ represents the expected service time from MEDEVAC base $i$ to serve casualty event $j$ and land at the nearest MTF. We use these basis functions as well as their interaction terms in order to approximate the value function.

### 3.3 Simulation

To compare the performance of the ADP against the myopic policy, we simulate multiple trajectories and compare the performance of the two policies. For the myopic policy, MEDEVAC requests are served with decreasing order of priority with first-in-first-out for like priorities. The flow chart for the model is shown in Figure 2.

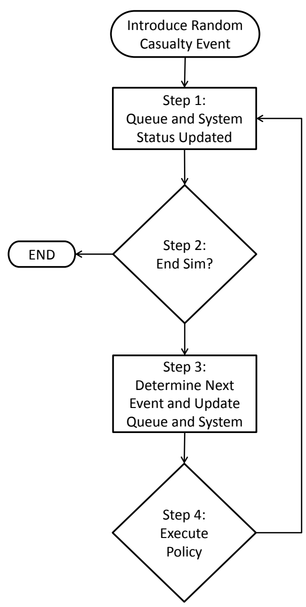



**Figure 2. Simulation Flow Chart**

The simulation is initiated by randomly generating a casualty event. A Hawkes spatial generation process is used for casualty event generation. The Hawkes process models situations where subsequent events are likely to occur in close proximity to the first event Kroese & Botev (2013). This distribution occurs according to a Poisson distribution whichwell models real-world casualty events. In Step 1, the queue and system status are updated. In Step 2, if termination time has not been met, the loop is continued. In Step 3, we stochastically determine if a casualty event has occurred (based on $\lambda$) or if a MEDEVAC event has occurred based on travel times and the stochastic elements. In Step 4, the system, total discounted reward, and queue are updated, the desired policy (ADP or myopic) is executed and the simulation returns to step 1. Step 4 incorporates the following logic check. If all MEDEVACs are *idle* and there is one or more casualty event waiting in the queue, the simulation is terminated and returns a result of "$DidNotFinish$"; this prevents premature convergence to a very poor policy.

# IV. Computational Results

In this chapter, we present a notional scenario to which we apply the ADP solution methods proposed in the previous chapter, and computational experiments are conducted using a scenario to obtain insights regarding solution quality and computational effort. We examine different features of the ADP algorithm and different features of the MEDEVAC dispatching problem to gain further insights regarding the performance of the proposed solution method. For the computational experiments, we use a dual Intel Xeon E5-2650v2 workstation with 192 GB of RAM and MATLAB's parallel computing toolbox.

## 4.1 Notional Scenario

We present a scenario in which a coalition of allied countries perform peacekeeping operations in response to islamic state militants in northern Syria. The locations for MEDEVAC bases are likely key military tactical sites. Casualty collection centers are selected and weighted by projected enemy locations. Figure 3 shows the 26 casualty cluster centers, five MEDEVAC locations, and two MTFs. Steady state and high operations tempo are assumed as is a baseline casualty event arrival rate of $\lambda = \frac{1}{60}$, representing an average casualty event arrival rate of one event per hour. Any MEDEVAC is allowed to service any casualty event and casualty events do not need to be served as soon as they arrive. Equal proportions of *urgent* and *priority* class casualty event arrivals are assumed. *Routine* events are not considered due to the high operational tempo, which is likely given intense combat scenarios where these *routine* events would be serviced by CASEVAC or ground MEDEVAC. The reward function utilizes weights of $\Delta_1$=10 and $\Delta_2$=2, which rewards *urgent* much greater than *priority* casualty events.

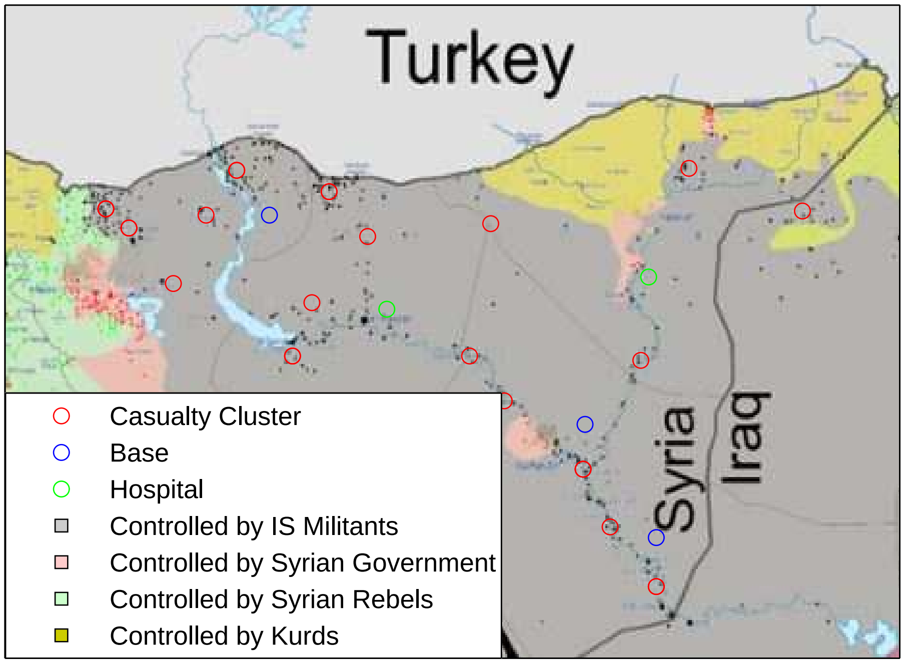



**Figure 3. Notional Scenario Disposition**

### 4.2 Experimental Design

For the experimental design, we focus on the problem parameter $\lambda$ and the ADP algorithmic parameters $\mathscr{G}$ and $\mathscr{H}$. The ADP parameters are set based on our initial experiences with implementing the model. An aggressive smoothing function is used because, without it, we obtain cyclical results from the policy improvement steps shown in Figure 4. With smoothing we rapidly approach the highest quality solution. We observe that any computation time beyond 20 policy iterations obtains little policy improvement. Similarly, after 20,000 samples for $\mathscr{H}$ we observed no improvement in performance. To compare solution quality and computation time we examine $\mathscr{G}$={5, 10, 20} and $\mathscr{H}$={5,000, 10,000, 20,000} at parameter levels $\lambda = \frac{1}{30}$, $\frac{1}{60}$ and $\frac{1}{120}$. The casualty arrival rates were chosen based on high operations tempo for $\lambda = \frac{1}{30}$ and $\frac{1}{60}$, and the ADP no longer out performs the myopic policy around $\lambda = \frac{1}{120}$.

A $3^3$ Factorial design is used to examine parameters in order to gain the most fidelity for the full design region. Table 3 shows factor levels.

**Table 3. Experimental Design**

| $\lambda$ | $\mathscr{G}$ | $\mathscr{H}$ |
|---|---|---|
| $\frac{1}{30}$ | 5 | 5,000 |
| $\frac{1}{60}$ | 10 | 10,000 |
| $\frac{1}{120}$ | 20 | 20,000 |

### 4.3 Experimental Design Results

Table 4 reports the results from the experimental design. The best performing features are highlighted for $\lambda = \frac{1}{30}$ and $\frac{1}{60}$. At $\lambda = \frac{1}{120}$ multiple features are within a 95% confidence interval. The general trend for the first two casualty arrival rates is 10 policy improvement loops and at least 10,000 policy evaluation loops. Computation time scales closely with the product $\mathscr{G} \cdot \mathscr{H}$. In order to effectively evaluate a policy, the ADP requires a minimum of 10,000 policy evaluation loops. The ADP converges quickly to the highest performing policy and policy improvement loops greater than 10 do not improve the ADP performance. As $\lambda$ increases, the ADP increasingly outperforms the myopic policy for all ADP algorithmic changes.

**Table 4. Experimental Design Results**

| Outter $\mathscr{G}$ | Inner $\mathscr{H}$ | $\frac{1}{\lambda}$ | Improvement | *Urgent* Wait Time in Sec. | *Priority* Wait Time in Sec. | MEDEVAC Busy % | Computation Time in Sec. |
|---|---|---|---|---|---|---|---|
| 5 | 5000 | 30 | 48.82 ± 0.92% | 182.5 | 398.0 | 90.7% | 18.8 |
| 10 | 5000 | 30 | 51.67 ± 0.96% | 186.8 | 388.2 | 90.6% | 37.3 |
| 20 | 5000 | 30 | 52.75 ± 0.92% | 190.8 | 377.4 | 90.5% | 74.7 |
| 5 | 10000 | 30 | 57.80 ± 0.93% | 187.9 | 375.0 | 90.4% | 37.5 |
| 10 | 10000 | 30 | 61.93 ± 0.86% | 191.7 | 369.3 | 88.7% | 74.8 |
| 20 | 10000 | 30 | 50.39 ± 0.94% | 186.7 | 386.5 | 90.6% | 149.8 |
| 5 | 20000 | 30 | 51.42 ± 0.89% | 177.8 | 397.2 | 90.6% | 74.7 |
| 10 | 20000 | 30 | 59.63 ± 0.91% | 190.5 | 369.7 | 89.6% | 150.2 |
| 20 | 20000 | 30 | 53.65 ± 0.90% | 192.0 | 373.5 | 90.6% | 300.1 |
| 5 | 5000 | 60 | 24.22 ± 0.99% | 170.0 | 234.3 | 84.1% | 19.1 |
| 10 | 5000 | 60 | 24.93 ± 1.07% | 170.4 | 230.1 | 83.0% | 37.1 |
| 20 | 5000 | 60 | 26.79 ± 1.04% | 177.1 | 265.4 | 80.1% | 74.7 |
| 5 | 10000 | 60 | 26.84 ± 1.05% | 166.7 | 232.8 | 83.2% | 37.8 |
| 10 | 10000 | 60 | 26.98 ± 1.09% | 168.2 | 232.0 | 82.4% | 74.9 |
| 20 | 10000 | 60 | 27.10 ± 1.00% | 167.9 | 230.8 | 82.0% | 150.5 |
| 5 | 20000 | 60 | 25.25 ± 0.99% | 167.0 | 240.4 | 84.2% | 74.9 |
| 10 | 20000 | 60 | 30.80 ± 1.07% | 154.0 | 396.5 | 81.0% | 149.9 |
| 20 | 20000 | 60 | 28.22 ± 1.01% | 165.6 | 237.8 | 81.4% | 300.1 |
| 5 | 5000 | 120 | 1.13 ± 1.43% | 115.2 | 120.3 | 34.8% | 19.1 |
| 10 | 5000 | 120 | -5.25 ± 1.57% | 121.5 | 132.3 | 36.4% | 37.5 |
| 20 | 5000 | 120 | -4.88 ± 1.55% | 117.9 | 127.0 | 33.8% | 75.4 |
| 5 | 10000 | 120 | -0.60 ± 1.47% | 116.6 | 121.7 | 35.0% | 37.7 |
| 10 | 10000 | 120 | -0.02 ± 1.53% | 113.9 | 121.0 | 32.3% | 74.6 |
| 20 | 10000 | 120 | -3.05 ± 1.53% | 119.3 | 145.4 | 32.4% | 150.1 |
| 5 | 20000 | 120 | 0.65 ± 1.55% | 113.3 | 117.5 | 32.9% | 75.2 |
| 10 | 20000 | 120 | -0.74 ± 1.50% | 117.2 | 123.7 | 35.4% | 149.6 |
| 20 | 20000 | 120 | -2.25 ± 1.52% | 116.2 | 129.2 | 32.7% | 300.6 |

## 4.4 Baseline Scenario Results and Analysis

The baseline scenario uses the parameters in Table 5. We use the best quality ADP settings for $\mathscr{G}$ and $\mathscr{H}$

**Table 5. Baseline Scenario Parameters**

| Problem Features | | | ADP Algorithm Features | | |
|---|---|---|---|---|---|
| Parameter | Description | Setting | Parameter | Description | Setting |
| $\lambda$ | Casualty Arrival Rate | 60 | $\phi$ | Basis Function | 3rd Order |
| $\Delta_1$ | Weight for *urgent* event | 10 | $\mathscr{G}$ | Policy Improvement Loop | 10 |
| $\Delta_2$ | Weight for *priority* event | 2 | $\mathscr{H}$ | Policy Evaluation Loop | 20000 |

Table 6 shows the performance of the ADP against the myopic policy. Based on a 95% confidence interval, the third order ADP outperforms the myopic policy by 30%. The average casualty event service times for the two casualty classification levels, and the average proportion of time that the MEDEVACs are *busy* (which includes time spent traveling back to base even if they are not actively serving a casualty event) also outperform the myopic policy. The ADP policy focuses on servicing *urgent* casualty events first, as noted by shorter wait times; moreover, MEDEVACs are being utilized more efficiently as shown by the lower average busy percentage. Utilizing a set of third order basis functions, achieves the best ADP performance. Use of a set of fourth order basis functions caused problems with computational singularity while calculating the LSTD regression Equation 11, so we did not experiment with higher order basis functions. There are no statistical differences in computational times between different orders of $\phi$. We use $\phi, \phi^2$, and $\phi^3$ for the rest of the experiments.

**Table 6. ADP Baseline Performance**

| Policy | Number of Basis Functions | Improvement Over Myopic | *urgent* service time (min.) | *priority* service time (min.) | MEDEVAC busy |
|---|---|---|---|---|---|
| Myopic | - | - | 285.4 | 286.7 | 89.3% |
| ADP, 1st Order | 160 | $17.7 \pm 1.4\%$ | 167.0 | 299.3 | 87.8% |
| ADP, 2nd Order | 320 | $26.0 \pm 1.5\%$ | 165.5 | 248.9 | 82.3% |
| ADP, 3rd Order | 480 | $30.8 \pm 1.4\%$ | 154.0 | 235.4 | 81.0% |

In Figure 4 we observe diminishing returns for policy improvement. Moreover, without the smoothing function, we fail to obtain a higher quality solution. We note this behavior across all experimental levels.

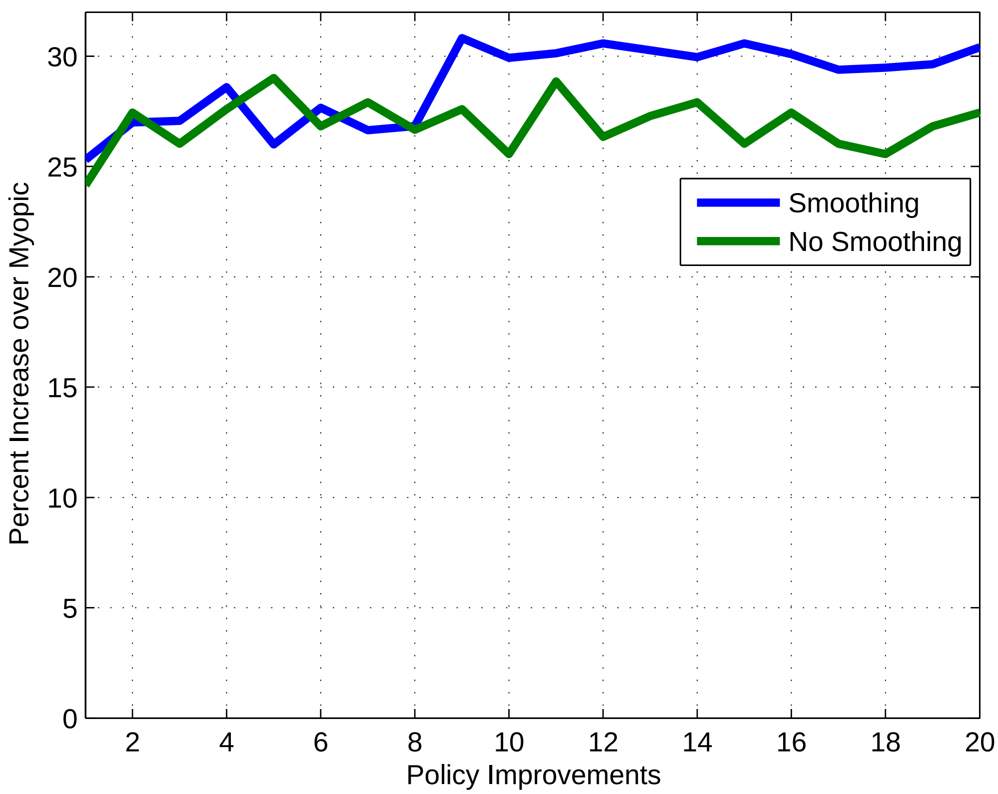



**Figure 4. Smoothing vs Non-Smoothing ADP Performance**

Examination of the basis functions from the best performing third order ADP results provides the following insight. The basis function $\phi_5$, which captures the interaction between MEDEVAC $i$ and the priority of the casualty event it is serving, has the largest impact on the policy. The number and proximity of casualty clusters to a MEDEVAC base show an increase in magnitude for their respective basis function coefficients ($\theta$). Moreover, MEDEVACs co-located with the MTF show a similar increase in magnitude, despite being further away from large groupings of casualty clusters. Interestingly, $\phi_1$, shows low statistical significance during regression and also has small magnitudes for its coefficients for all MEDEVACs. This is because the ADP policy does not seek to penalize MEDEVACs for being idle, which could force them to otherwise launch on low priority or far away casualty events.

Examination of the best ADP policy indicates that it is best to launch MEDEVACs which are close to casualty events and co-located with hospitals only on high priority casualty events. It is difficult to determine the dynamics of when, and for how long, to hold a MEDEVAC in reserve before launching because of the sheer dimensionality of the system.

We next examine the performance of the ADP against different casualty rates by adjusting $\lambda$. Figure 5 shows the performance of the ADP against the myopic policy.

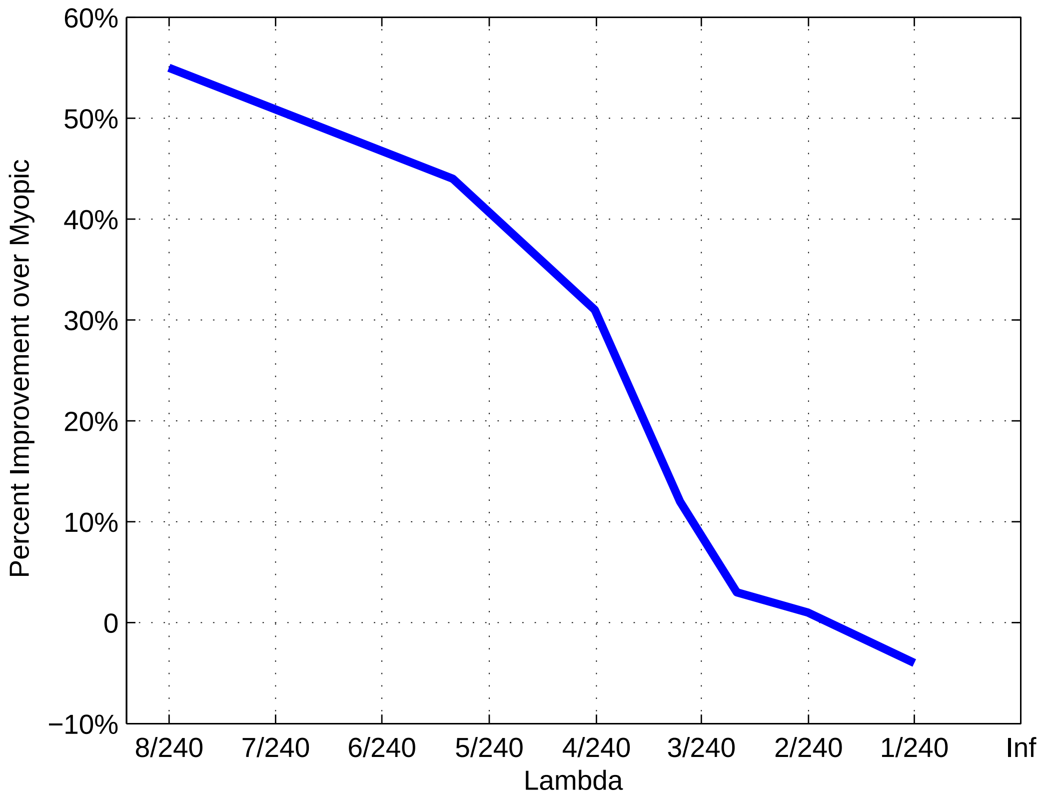



**Figure 5. Percent Improvement Over Myopic**

As $\lambda$ decreases, the frequency with which casualties arrive decreases and the ADP policies no longer outperform the myopic policy. As casualty events arrive at a slower rate, the utility of holding MEDEVAC aircraft in reserve is diminished. At a casualty event arrival rate of one every two hours (i.e. $\lambda = \frac{1}{240}$ we observe a reduction in performance of the ADP compared to the myopic policy, showing the limitation of the set of basis functions and their resulting value function approximations. Keneally *et al.* (2014) show that as $\lambda$ goes to extremes, the difference in performance between optimal and myopic policies becomes negligible.

The impact of changing the proportion of casualties from all *urgent* to all *priority* casualty events is shown in Figure 6. We observe a noted improvement over the myopic policy. This indicates the ADP policy is efficiently managing resources by not immediately sending the nearest available MEDEVAC, but rather waiting and/or sending a MEDEVAC that is further away but is also not close to a high rate casualty cluster area.

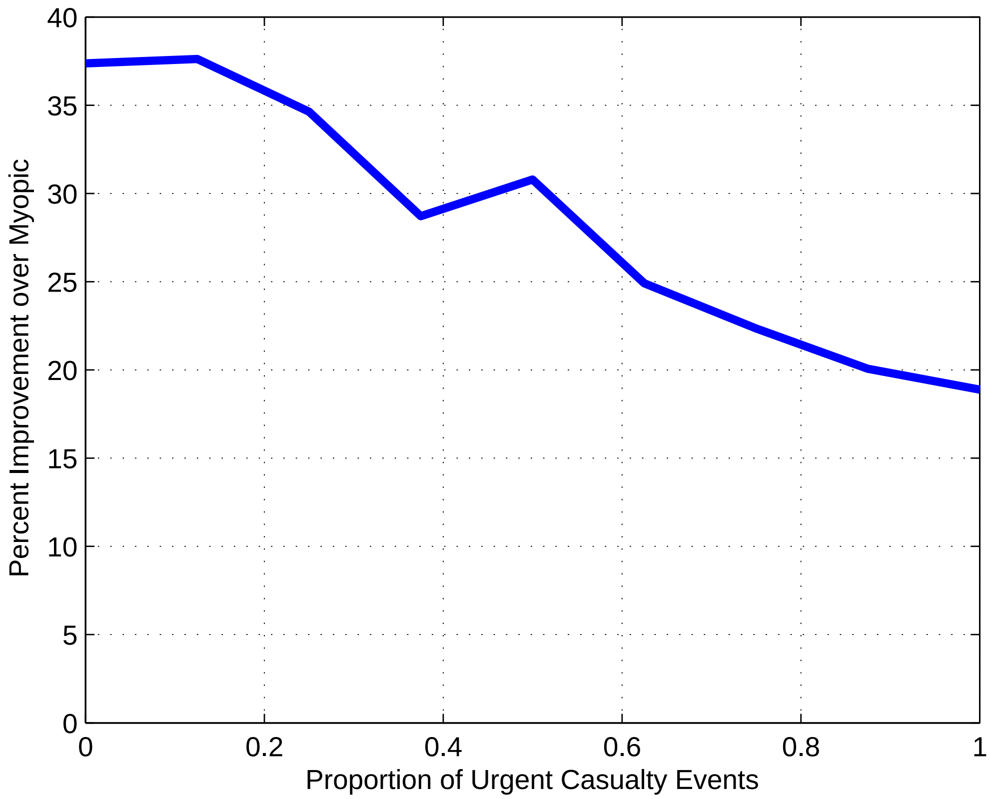



**Figure 6. Performance of Changing Proportion**

We also examine the impact of using a different reward function. Instead of a simple exponential decay, a monotonically decreasing arctan function is utilized. The function slowly decays before reaching a threshold in proximity of the requirements outlined by Department of the Army (2007). Figure 7 shows the comparison of the exponential decay versus the slower decaying arctan function.

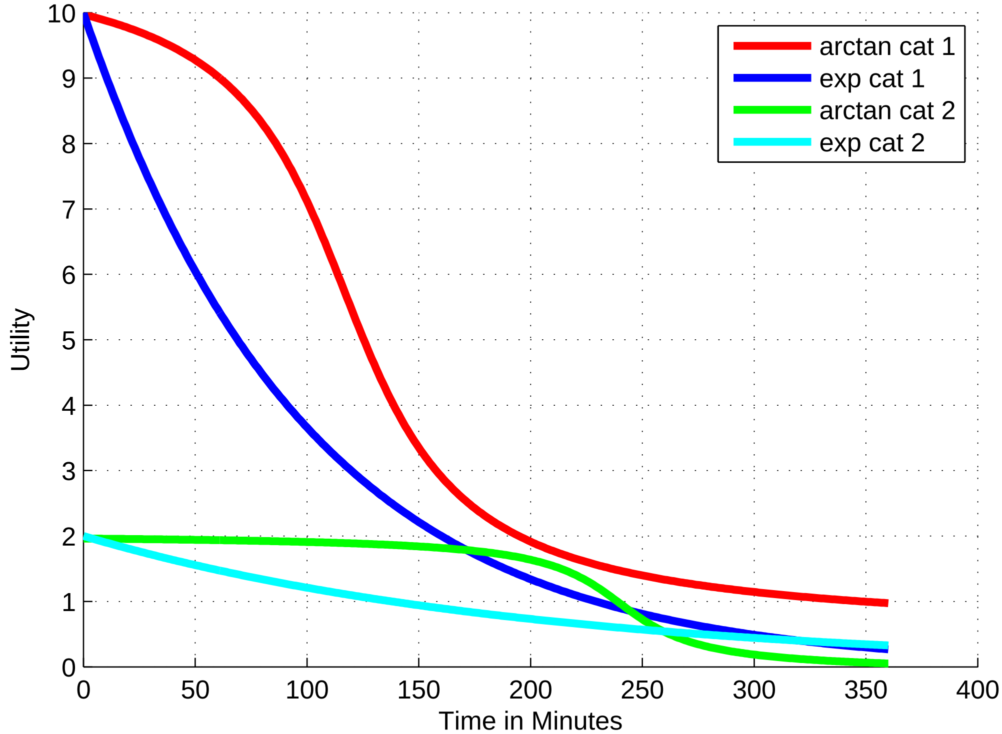



**Figure 7. Reward Functions**

Table 7 shows the performance of the ADP policy versus the myopic policy with both reward functions in the baseline scenario. We observe nearly identical performance between the two reward functions with overlapping 95% confidence intervals. These results indicate the ADP policy is robust to changes in the reward function.

**Table 7. Reward Function Comparison**

| Reward Function | Performance |
|---|---|
| Exponential | 30.8% $\pm$ 1.40% |
| Arctan | 31.0% $\pm$ 1.48% |

Lastly, there are experimental rotary wing aircraft which could potentially be put into service and which can travel significantly faster than the UH-60 A/L Blackhawk. To examine the impact of these aircraft, the maximum speed with which the MEDEVAC aircraft can travel is adjusted, and stochastic parameters remain constant. Table

8 shows the percent increase in average airspeed over the UH-60 A/L.

**Table 8. Performance vs increasing Airspeed**

| | Airspeed Increase vs UH-60 | | | |
|---|---|---|---|---|
| | Baseline | 25% | 50% | 75% |
| Myopic Improvement (Against Baseline Myopic) | - | 50.05% | 89.45% | 126.12% |
| Urgent Service Time (min.) | 286.1 | 168.5 | 104.8 | 77.2 |
| Priority Service Time (min.) | 284.4 | 166.8 | 104.1 | 76.8 |
| MEDEVAC Busy % | 89.3% | 78.6% | 61.6% | 47.8% |
| ADP Improvement (Against Baseline Myopic) | 30.78% | 68.32% | 102.81% | 127.12% |
| ADP Improvement (Against Like Myopic) | 30.78% | 12.17% | 7.05% | 0.44% |
| Urgent Service Time (min.) | 154.0 | 116.8 | 90.8 | 74.8 |
| Priority Service Time (min.) | 396.5 | 140.3 | 103.4 | 81.5 |
| MEDEVAC Busy % | 70.2% | 69.6% | 56.6% | 46.9% |

It is reasonable to assume newer rotary wing designs can increase airpseed 25%-50% compared the the UH-60 A/L. This increase in speed has a significant impact in the overall performance for both ADP and myopic policies. We see diminishing returns for the ADP compared to the myopic policy as the airspeed is increased. Despite this, decision makers can still benefit from high quality dispatching solutions.

When we examine computational effort more closely, we find the ADP solution only requires about 150 seconds for the baseline scenario. The largest computational effort came from running the simulation. For each event in the simulation, the best action must be determined, as indicated by Equation 9. In addition to these calculations, we require 500 runs in order to achieve our desired confidence interval. The computation times for the simulation using the myopic and ADP policies are shown in Table 9.

**Table 9. Simulation run times**

| Policy | Computation Time (sec.) | Runs |
|---|---|---|
| Myopic | 34.625 | 500 |
| ADP | 884.93 | 500 |

The computation time of calculating Equation 9 is about 25 times that of the hard-coded myopic policy. Despite this burden, it is still possible to compute the ADP policy and run the simulation in under 20 minutes. These results are promising, as the model could readily be adapted and applied to current operations to yield timely results.

# V. Conclusions

This thesis examines the MEDEVAC dispatching problem. The intent of the research is to determine policies that increase the survivability of battlefield casualties. Development of an MDP model of the MEDEVAC dispatching problem enables examination of many different scenarios. Solving the MDP requires the use of ADP. By using the *post-decision* state and approximating the value of being that state, our model became computationally tractable. To examine the performance of policies produced by our model, we created a scenario and simulated the outcome of the established policies.

The ADP policy was able to increase overall utility by 30% compared to the myopic policy in our baseline scenario. Additionally, MEDEVAC *busy* time was decreased by 9%, indicating more efficient use of MEDEVAC aircraft. The basis function coefficients revealed MEDEVAC aircraft in close proximity to higher probability casualty clusters were more valuable than aircraft based further away. This is an intuitive result. These higher value MEDEVACs should likely not be dispatched for low priority casualty events while the lower value MEDEVACs may be dispatched instead. The ADP policy was able capture the overall time it would take for any MEDEVAC to service any casualty event. This is important, as a MEDEVAC may possibly become available which could service a casualty event faster than an *idle* MEDEVAC that is further away. Maximum speed of the aircraft has the largest impact on performance. Results indicate a 25% increase in speed increases utility by 50%. Even with the performance increase from speed, the ADP policy still provides increased utility compared to the myopic policy.

This model and its results are beneficial to military planners and decisions makers. Military planners can use this model to compare policies as well as evaluate different potential MEDEVAC station locations in order to maximize performance.

Decision makers can use current military intelligence and operational experience to identify areas in which casualties are likely to occur. Once these areas are identified, decisions makers can make informed decisions about the value of each MEDEVAC and maximize the utilization of their resources.

Results also indicate the criticality of MEDEVAC travel speed. Military planners and acquisitions (those responsible for implementing new technology into the military) can use this model to examine the impact of capacity compared to speed. This information can be used for future design and development for a replacement to the UH-60 A/L. Perhaps a mix of large capacity UH-60 A/Ls and a new-lower capacity, high-speed design would improve overall casualty survivability.

The model does not take into account dynamic repositioning or dispatching. Aircraft are required to return to base before they become available. In many situations, decision makers will immediately dispatch a MEDEVAC who dropped off a casualty event at an MTF, but has bit yet returned to their original base, to a new casualty event. This possibility was not considered in this thesis as crew limitations do not always allow this decision to be feasible.

Dynamic in-flight rerouting is not considered. If a MEDEVAC has capacity to take on additional casualties and a casualty event occurs in close proximity, it may be worth sacrificing time for the casualties on-board in order to reduce the service time for the new casualty event. Communication limitations as well as uncertainty about specific casualty events (e.g, the actual condition of the casualty event onboard versus the survivability function) make this a complex decision. However, for low priority casualty events, dynamic rerouting would likely have significantly less negative impact on those events and have a large positive impact on the casualty event which would be served faster.

Future extensions to this model could include dynamic routing and rerouting, comparing different aircraft types, and comparing MEDEVAC placement (i.g, multiple aircraft at a base and some bases empty). For a long range planning tool, researchers should focus on the impact of MEDEVAC placement and aircraft designs. These results could provide valuable insight for military planners.

Implementing the ADP policy into active operations is an incredibly difficult proposition. The myopic policy is often used because it is simple to implement and performs fairly well as long as casualty events arrive at a frequency less than two hours. The important point for decisions makers to garner is understanding the value of specific MEDEVACs and how to utilize them efficiently.

# VI. Appendix

## 6.1 Appendix A

| | |
|---|---|
| $\lambda$ | Arrival rate of casualty events |
| $s$ | The current state of the system |
| $\tau$ | Current system time |
| $\epsilon$ | Current event |
| $a$ | Number of MEDEVAC Units |
| $m_i$ | Information about the $i$th MEDEVAC |
| $b$ | Size of the casualty event queue |
| $q_j$ | Information about the casualty event in the $j$th position of the queue |
| $\sigma_i$ | Status of the MEDEVAC |
| $d_i$ | Expected time to complete current movement |
| $t_i$ | Start time of current MEDEVAC movement |
| $\delta_j$ | Status of the casualty event |
| $l_j$ | Location of the casaulty event |
| $\zeta_j$ | Time the casualty event arrived in the system |
| $\eta_j$ | Priority of the casualty event |
| $x_k$ | Decision after the $k$th event |
| $\omega$ | Stochastic process for state transition |
| $\Delta$ | Utility multiplier for casualty event priorities |
| $\alpha$ | Harmonic step size parameter |
| $d^r_{i,k}$ | Expected time for MEDEVAC $i$ to return to base after event $k$ |
| $d^*_{i,j}$ | Expected time from base $i$ to pickup and drop off casualty event $j$ |
| $\mathscr{G}$ | The number of Policy Improvement Loops |
| $\mathscr{H}$ | The number of states sampled per Policy Evaluation Loop |
| $\mathscr{M}$ | The set of *idle* MEDEVAC aircraft |
| $\mathscr{Q}$ | The set of *idle* casualty events |
| $\mathscr{X}$ | The set of feasible actions |

**Table 10. Parameters**

## 6.2 Appendix B

# An Approximate Dynamic Programming Model for MEDEVAC Dispatching


CPT Aaron J. Rettke
Advisor:
Lt Col Matthew J. Robbins, Ph.D.
Reader:
LTC Brian J. Lunday, Ph.D.
Department of Operational Sciences (ENS)


### Research Objective

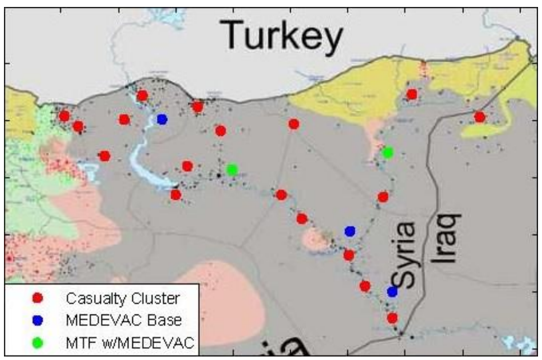
**Inform development or modification to MEDEVAC Tactics Techniques and Procedures (TTP) to increase casualty survivability.**

**Currently, 90% of casualties survive on the battlefield. This is achieved by using a myopic policy, which sends the nearest available MEDEVAC. We seek MEDEVAC dispatching policies that increase casualty survivability.**

**A Markov Decision Process (MDP) model is developed to examine the MEDEVAC dispatching problem. An Approximate Dynamic Program (ADP) solution methodology is employed to obtain high-quality policies.**

### Notional Scenario

**We create a notional scenario in which coalition forces respond to Islamic State militants in northern Syria.**




**In this scenario, 5 MEDEVACs and 2 Medical Treatment Facilities (MTF) are emplaced as depicted. Casualty events occur at a rate of one every hour, indicating a high operations tempo. MEDEVAC units must return to their original base after servicing a casualty event to represent maintenance and crew limitations.**

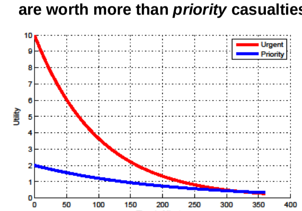
### Solution Methodology

**A Least Squares Approximate Policy iteration (LSAPI) algorithm is developed. We create basis functions, φ, to approximate the value of being in a particular *post-decision* state. θ represents the basis function weights, which we use to calculate value of a specific *post-decision* state.**

Step 1: Initialize $\theta$
Step 2: for $g = 1$ to $\mathcal{G}$ (Policy Improvment Loop)
Step 3: for h = 1 to $\mathcal{H}$ (Policy Evaluation Loop/Sample)
Step 4: Simulate a random post-decision state, $S^x_{k-1,h}$
Step 5: Record $\phi(S^x_{k-1,h})$
Step 6: Simulate the state transition for the next event to get $S^x_{k,h}$
Step 7: Determine the decision, $x_{ij} = X^{\pi}(S_{k,h} \mid \theta)$
Step 8: Record $r(S_{k,h})$
Step 9: Record $\phi(S^x_{k,h})$, the observation $\mathbb{E}[\phi(S^x_{k,h}) \mid \phi(S^x_{k-1,h})]$
Step 10: End
Step 11: Update $\theta$ using LSTD and a harmonic smoothing function
Step 12: End

### Survivability Function

**An exponentially decreasing utility function is used to model casualty degradation over time. *Urgent* casualties are worth more than *priority* casualties.**

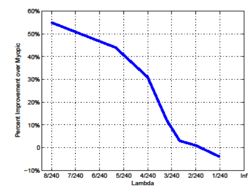



### Results

**Our baseline results show a 30% increase in performance over a myopic (send the nearest available unit) policy. The ADP policy achieves this by prioritizing *Urgent* casualties, as well as being more efficient when utilizing MEDEVACs, as indicated by the busy percentage.**

| Policy | Improvement Over Myopic | Urgent service time (min.) | Priority service time (min.) | MEDEVAC Busy |
|---|---|---|---|---|
| Myopic | - | 285.4 | 286.7 | 89.3% |
| ADP | 30.8±1.4% | 154 | 235.4 | 81.0% |

**Adjusting the casualty arrival rate shows the ADP policy does better than the myopic policy until a casualty event arrival rate of one every two hours.**




| | Airspeed Increase vs UH-60 A/L: Baseline | 25% | 50% | 75% |
|---|---|---|---|---|
| Myopic Improvement (Against baseline myopic) | - | 50.1% | 89.5% | 126.1% |
| Urgent Service Time (min.) | 286.1 | 168.5 | 104.8 | 77.2 |
| Priority Service Time (min.) | 284.4 | 166.8 | 104.1 | 76.8 |
| MEDEVAC Busy % | 89.3% | 78.6% | 61.6% | 47.8% |
| ADP Improvement (Against baseline myopic) | 30.8% | 68.3% | 102.8% | 127.1% |
| ADP Improvement (Against like myopic) | 30.8% | 12.2% | 7.0% | 4.0% |
| Urgent Service Time (min.) | 154.0 | 116.8 | 90.8 | 74.8 |
| Priority Service Time (min.) | 365.5 | 140.3 | 103.4 | 81.5 |
| MEDEVAC Busy % | 70.2% | 69.6% | 56.6% | 46.9% |

**The table shows the impact of increasing aircraft airspeed. Increasing airspeed has the largest impact on the system. The ADP policy remains superior to the myopic policy.**

**When examining the θ coefficients we note that MEDEVAC units that are close in proximity to multiple casualty clusters have more magnitude and effect on the system. This matches our intuition, which is that these MEDEVACs need to be used only in response to *Urgent* casualties, while other, less weighted MEDEVACs can be dispatched for lower *Priority* events.**

### Implications

**Planners can use this model to easily examine MEDEVAC dispatching in a multitude of scenarios.**

**Military operations can use this as supporting evidence to understand the importance of identifying high threat areas and their impact on MEDEVAC dispatching decisions and TTPs.**

**Military acquisitions and decision makers can use this model to examine the impact of increasing aircraft airspeed and consider its implications when considering replacements for the current MEDEVAC aircraft, the UH-60 A/L Blackhawk.**

Department of Operational Sciences

# Bibliography


Alsalloum, Othman Ibraheem, & Rand, Graham K. 2006. Extensions to emergency vehicle location models. *Computers & Operations Research*, **33**(9), 2725–2743.

Bandara, Damitha, Mayorga, Maria E, & McLay, Laura A. 2012. Optimal dispatching strategies for emergency vehicles to increase patient survivability. *International Journal of Operational Research*, **15**(2), 195–214.

Bastian, Nathaniel D. 2010. A robust, multi-criteria modeling approach for optimizing aeromedical evacuation asset emplacement. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, **7**(1), 5–23.

Batta, Rajan, Dolan, June M, & Krishnamurthy, Nirup N. 1989. The maximal expected covering location problem: Revisited. *Transportation Science*, **23**(4), 277–287.

Bradtke, Steven J, & Barto, Andrew G. 1996. Linear least-squares algorithms for temporal difference learning. *Machine Learning*, **22**(1-3), 33–57.

Carter, Grace M, Chaiken, Jan M, & Ignall, Edward. 1972. Response areas for two emergency units. *Operations Research*, **20**(3), 571–594.

Church, Richard, & ReVelle, Charles R. 1974. The maximal covering location problem. *Papers in Regional Science*, **32**(1), 101–118.

Department of the Army, United States. 2007. *Field Manual 4-02.2, Medical Evacuation*.

Eastridge, Brian J, Mabry, Robert L, Seguin, Peter, Cantrell, Joyce, Tops, Terrill, Uribe, Paul, Mallett, Olga, Zubko, Tamara, Oetjen-Gerdes, Lynne, Rasmussen, Todd E, *et al.* 2012. Death on the battlefield (2001–2011): implications for the future of combat casualty care. *Journal of Trauma and Acute Care Surgery*, **73**(6), S431–S437.

Erkut, Erhan, Ingolfsson, Armann, & Erdoğan, Güneş. 2008. Ambulance location for maximum survival. *Naval Research Logistics*, **55**(1), 42–58.

Feero, Stan, Hedges, Jerris R, Simmons, Erik, & Irwin, Lisa. 1995. Does out-of-hospital EMS time affect trauma survival? *The American Journal of Emergency Medicine*, **13**(2), 133–135.

Fulton, Lawrence V, Lasdon, Leon S, McDaniel, Reuben R, & Coppola, M Nicholas. 2010. Two-stage stochastic optimization for the allocation of medical assets in steady-state combat operations. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, **7**(2), 89–102.

Garrett, Michael X. 2013. *USCENTCOM Review of MEDEVAC Procedures in Afghanistan*. Tech. rept. United States Central Command.

Grannan, Benjamin, Bastian, Nathaniel, & McLay, Laura. 2014. A Maximum Expected Covering Problem for Locating and Dispatching Two Classes of Military Medical Evacuation Air Assets. *Optimization Letters*.

Hartenstein, Ingo. 2008. Medical Evacuation in Afghanistan: Lessons identified! Lessons Learned?

Howard, Ronald A. 1960. *Dynamic Programming and Markov Processes*. MIT Press, Cambridge, Massachusetts.

Keneally, Sean, Robbins, Matthew, & Lunday, Brian. 2014. A Markov Decision Process Model for the Optimal Dispatch of Military Medical Evacuation Assets. *Health Care Management Science*.

Kroese, Dirk P, & Botev, Zdravko I. 2013. Spatial Process Generation. *Lectures on Stochastic Geometry, Spatial Statistics and Random Fields*, **2**.

Lagoudakis, Michail G, & Parr, Ronald. 2003. Least-squares policy iteration. *The Journal of Machine Learning Research*, **4**, 1107–1149.

Lerner, E Brooke, & Moscati, Ronald M. 2001. The golden hour: scientific fact or medical urban legend? *Academic Emergency Medicine*, **8**(7), 758–760.

Ma, Jun, & Powell, Warren B. 2010. Convergence analysis of kernel-based on-policy approximate policy iteration algorithms for markov decision processes with continuous, multidimensional states and actions. *Princeton University*.

Maxwell, Matthew S, Restrepo, Mateo, Henderson, Shane G, & Topaloglu, Huseyin. 2010. Approximate dynamic programming for ambulance redeployment. *INFORMS Journal on Computing*, **22**(2), 266–281.

Mayorga, Maria E, Bandara, Damitha, & McLay, Laura A. 2013. Districting and dispatching policies for emergency medical service systems to improve patient survival. *IIE Transactions on Healthcare Systems Engineering*, **3**(1), 39–56.

McLay, Laura A, & Mayorga, Maria E. 2010. Evaluating emergency medical service performance measures. *Health Care Management Science*, **13**(2), 124–136.

Powell, Warren B. Perspectives of approximate dynamic programming. *Annals of Operations Research*, **13**.

Powell, Warren B. 2009. What you should know about approximate dynamic programming. *Naval Research Logistics*, **56**(3), 239–249.

ReVelle, Charles, & Hogan, Kathleen. 1989. The maximum availability location problem. *Transportation Science*, **23**(3), 192–200.

Scott, Warren, Powell, Warren, & Moazeni, Somayeh. 2014. *Least Squares Policy Iteration with Instrumental Variables vs. Direct Policy Search: Comparison Against Optimal Benchmarks Using Energy Storage*. Tech. rept. Princeton.

Silva, Francisco, & Serra, Daniel. 2008. Locating emergency services with different priorities: the priority queuing covering location problem. *Journal of the Operational Research Society*, **59**(9), 1229–1238.

Spurgeon, Neel. 1991. *Medical Support of the U.S. Army in Vietnam*. Department of the Army.

Zeto, John, Yamada, Wade, & Collins, Gale. 2006. Optimizing the Emplacement of Scarce Resources to Maximize the Expected Coverage of a Geographically Variant Demand Function. *Center for Army Analysis*.

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704–0188*



**1. REPORT DATE** *(DD–MM–YYYY)*
26–03–2015

**2. REPORT TYPE**
Master's Thesis

**3. DATES COVERED** *(From — To)*
Oct 2013 — Mar 2015

**4. TITLE AND SUBTITLE**
An Approximate Dynamic Programming Mode
for Optimal MEDEVAC Dispatching

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
Rettke, Aaron J.,Captain, USA

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/ENS)
2950 Hobson Way
WPAFB OH 45433-7765

**8. PERFORMING ORGANIZATION REPORT NUMBER**
AFIT-ENS-MS-15-M-115

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Intentionally Left Blank

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION / AVAILABILITY STATEMENT**
Distribution Statement A.
Approved for Public Release; distribution unlimited.

**13. SUPPLEMENTARY NOTES**


**14. ABSTRACT**
We develop a Markov decision process (MDP) model to examine military medical evacuation (MEDEVAC) dispatch policies. To solve our MDP, we apply an approximate dynamic programming (ADP) technique. The problem of deciding which aeromedical asset to dispatch to which service request is complicated by the service locations and the priority class of each casualty event. We assume requests for MEDEVAC arrive sequentially, with the location and the priority of each casualty known upon initiation of the request. The proposed model finds a high quality dispatching policy which outperforms the traditional myopic policy of sending the nearest available unit. Utility is gained by servicing casualties based on both their priority and the actual time until a casualty arrives at a medical treatment facility (MTF). The model is solved using approximate policy iteration (API) and least squares temporal difference (LSTD). Computational examples are used to investigate dispatch policies for a scenario set in northern Syria. Results indicate that a myopic policy is not always the best policy to use for quickly dispatching MEDEVAC units, and insight is gained into the value of specific MEDEVAC locations.

**15. SUBJECT TERMS**
Emergency Medical Dispatch, medical evacuation (MEDEVAC), Markov decision processes, approximate dynamic programming, approximate policy iteration, least squares temporal difference

| 16. SECURITY CLASSIFICATION OF: a. REPORT | b. ABSTRACT | c. THIS PAGE | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES |
|---|---|---|---|---|
| U | U | U | UU | 51 |

**19a. NAME OF RESPONSIBLE PERSON**
Matthew J. Robbins, AFIT/ENS

**19b. TELEPHONE NUMBER** *(include area code)*
(937) 255-3636, x4539;matthew.robbins@afit.edu

**Standard Form 298 (Rev. 8–98)**
**Prescribed by ANSI Std. Z39.18**